

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Bujard, Hermann
Gossen, Manfred
Salfeld, Jochen G.
Voss, Jeffrey W.

(ii) TITLE OF INVENTION: Animals Transgenic for a Tetracycline-
Controlled Transcriptional Transactivator

(iii) NUMBER OF SEQUENCES: 10

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Lahive & Cockfield
(B) STREET: 60 State Street
(C) CITY: Boston
(D) STATE: Massachusetts
(E) COUNTRY: USA
(F) ZIP: 02109-1875

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: ASCII text

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 08/383,754
(B) FILING DAE: 14-JUN-1994

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 08/076,327
(B) FILING DAE: 14-JUN-1993

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: DeConti, Giulio A., Jr.
(B) REGISTRATION NUMBER: 31,503
(C) REFERENCE/DOCKET NUMBER: BBI-013CP2

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (617) 227-7400
(B) TELEFAX: (617) 227-5941

INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1008 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

10

(A) ORGANISM: Herpes Simplex Virus
(B) STRAIN: K12, KOS

(vii) IMMEDIATE SOURCE

(B) CLONE: tTA transactivator

15

(ix) FEATURE:

(A) NAME/KEY: exon
(B) LOCATION: 1..1008

(ix) FEATURE:

20

(A) NAME/KEY: mRNA
(B) LOCATION: 1..1008

(ix) FEATURE:

25

(A) NAME/KEY: misc. binding
(B) LOCATION: 1..207

(ix) FEATURE:

(A) NAME/KEY: misc. binding
(B) LOCATION: 208..335

30

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 1..1005

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATG TCT AGA TTA GAT AAA AGT AAA GTG ATT AAC AGC GCA TTA GAG CTG	48
Met Ser Arg Leu Asp Lys Ser Lys Val Ile Asn Ser Ala Leu Glu Leu	
1 5 10 15	
CTT AAT GAG GTC GGA ATC GAA GGT TTA ACA ACC CGT AAA CTC GCC CAG	96
Leu Asn Glu Val Gly Ile Glu Gly Leu Thr Thr Arg Lys Leu Ala Gln	
20 25 30	
AAG CTA GGT GTA GAG CAG CCT ACA TTG TAT TGG CAT GTA AAA AAT AAG	144
Lys Leu Gly Val Glu Gln Pro Thr Leu Tyr Trp His Val Lys Asn Lys	
35 40 45	
CGG GCT TTG CTC GAC GCC TTA GCC ATT GAG ATG TTA GAT AGG CAC CAT	192
Arg Ala Leu Leu Asp Ala Leu Ala Ile Glu Met Leu Asp Arg His His	
50 55 60	
ACT CAC TTT TGC CCT TTA GAA GGG GAA AGC TGG CAA GAT TTT TTA CGT	240
Thr His Phe Cys Pro Leu Glu Gly Glu Ser Trp Gln Asp Phe Leu Arg	
65 70 75 80	
AAT AAG GCT AAA AGT TTT AGA TGT GCT TTA CTA AGT CAT CGC GAT GGA	288
Asn Lys Ala Lys Ser Phe Arg Cys Ala Leu Leu Ser His Arg Asp Gly	
85 90 95	

	GCA AAA GTA CAT TTA GGT ACA CGG CCT ACA GAA AAA CAG TAT GAA ACT	336
	Ala Lys Val His Leu Gly Thr Arg Pro Thr Glu Lys Gln Tyr Glu Thr	
	100 105 110	
5	CTC GAA AAT CAA TTA GCC TTT TTA TGC CAA CAA GGT TTT TCA CTA GAG	384
	Leu Glu Asn Gln Leu Ala Phe Leu Cys Gln Gln Gly Phe Ser Leu Glu	
	115 120 125	
10	AAT GCA TTA TAT GCA CTC AGC GCT GTG GGG CAT TTT ACT TTA GGT TGC	432
	Asn Ala Leu Tyr Ala Leu Ser Ala Val Gly His Phe Thr Leu Gly Cys	
	130 135 140	
15	GTA TTG GAA GAT CAA GAG CAT CAA GTC GCT AAA GAA GAA AGG GAA ACA	480
	Val Leu Glu Asp Gln Glu His Gln Val Ala Lys Glu Glu Arg Glu Thr	
	145 150 155 160	
20	CCT ACT ACT GAT AGT ATG CCG CCA TTA TTA CGA CAA GCT ATC GAA TTA	528
	Pro Thr Thr Asp Ser Met Pro Pro Leu Leu Arg Gln Ala Ile Glu Leu	
	165 170 175	
	TTT GAT CAC CAA GGT GCA GAG CCA GCC TTC TTA TTC GGC CTT GAA TTG	576
	Phe Asp His Gln Gly Ala Glu Pro Ala Phe Leu Phe Gly Leu Glu Leu	
	180 185 190	
25	ATC ATA TGC GGA TTA GAA AAA CAA CTT AAA TGT GAA AGT GGG TCC GCG	624
	Ile Ile Cys Gly Leu Glu Lys Gln Leu Lys Cys Glu Ser Gly Ser Ala	
	195 200 205	
30	TAC AGC CGC GCG CGT ACG AAA AAC AAT TAC GGG TCT ACC ATC GAG GGC	672
	Tyr Ser Arg Ala Arg Thr Lys Asn Asn Tyr Gly Ser Thr Ile Glu Gly	
	210 215 220	
35	CTG CTC GAT CTC CCG GAC GAC GAC GCC CCC GAA GAG GCG GGG CTG GCG	720
	Leu Leu Asp Leu Pro Asp Asp Ala Pro Glu Glu Ala Gly Leu Ala	
	225 230 235 240	
40	GCT CCG CGC CTG TCC TTT CTC CCC GCG GGA CAC ACG CGC AGA CTG TCG	768
	Ala Pro Arg Leu Ser Phe Leu Pro Ala Gly His Thr Arg Arg Leu Ser	
	245 250 255	
45	ACG GCC CCC CCG ACC GAT GTC AGC CTG GGG GAC GAG CTC CAC TTA GAC	816
	Thr Ala Pro Pro Thr Asp Val Ser Leu Gly Asp Glu Leu His Leu Asp	
	260 265 270	
	GGC GAG GAC GTG GCG ATG GCG CAT GCC GAC GCG CTA GAC GAT TTC GAT	864
	Gly Glu Asp Val Ala Met Ala His Ala Asp Ala Leu Asp Asp Phe Asp	
	275 280 285	
50	CTG GAC ATG TTG GGG GAC GGG GAT TCC CCG GGT CCG GGA TTT ACC CCC	912
	Leu Asp Met Leu Gly Asp Gly Asp Ser Pro Gly Pro Gly Phe Thr Pro	
	290 295 300	
55	CAC GAC TCC GCC CCC TAC GGC GCT CTG GAT ATG GCC GAC TTC GAG TTT	960
	His Asp Ser Ala Pro Tyr Gly Ala Leu Asp Met Ala Asp Phe Glu Phe	
	305 310 315 320	
	GAG CAG ATG TTT ACC GAT CCC CTT GGA ATT GAC GAG TAC GGT GGG TAG	1008
	Glu Gln Met Phe Thr Asp Pro Leu Gly Ile Asp Glu Tyr Gly Gly	

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 335 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

10 Met Ser Arg Leu Asp Lys Ser Lys Val Ile Asn Ser Ala Leu Glu Leu
 1 5 10 15
 Leu Asn Glu Val Gly Ile Glu Gly Leu Thr Thr Arg Lys Leu Ala Gln
 20 25 30
 15 Lys Leu Gly Val Glu Gln Pro Thr Leu Tyr Trp His Val Lys Asn Lys
 35 40 45
 Arg Ala Leu Leu Asp Ala Leu Ala Ile Glu Met Leu Asp Arg His His
 20 50 55 60
 Thr His Phe Cys Pro Leu Glu Gly Glu Ser Trp Gln Asp Phe Leu Arg
 65 70 75 80
 25 Asn Lys Ala Lys Ser Phe Arg Cys Ala Leu Leu Ser His Arg Asp Gly
 85 90 95
 Ala Lys Val His Leu Gly Thr Arg Pro Thr Glu Lys Gln Tyr Glu Thr
 100 105 110
 30 Leu Glu Asn Gln Leu Ala Phe Leu Cys Gln Gln Gly Phe Ser Leu Glu
 115 120 125
 Asn Ala Leu Tyr Ala Leu Ser Ala Val Gly His Phe Thr Leu Gly Cys
 35 130 135 140
 Val Leu Glu Asp Gln Glu His Gln Val Ala Lys Glu Glu Arg Glu Thr
 145 150 155 160
 40 Pro Thr Thr Asp Ser Met Pro Pro Leu Leu Arg Gln Ala Ile Glu Leu
 165 170 175
 Phe Asp His Gln Gly Ala Glu Pro Ala Phe Leu Phe Gly Leu Glu Leu
 180 185 190
 45 Ile Ile Cys Gly Leu Glu Lys Gln Leu Lys Cys Glu Ser Gly Ser Ala
 195 200 205
 Tyr Ser Arg Ala Arg Thr Lys Asn Asn Tyr Gly Ser Thr Ile Glu Gly
 50 210 215 220
 Leu Leu Asp Leu Pro Asp Asp Asp Ala Pro Glu Glu Ala Gly Leu Ala
 225 230 235 240
 55 Ala Pro Arg Leu Ser Phe Leu Pro Ala Gly His Thr Arg Arg Leu Ser

	245	250	255
	Thr Ala Pro Pro Thr Asp Val Ser Leu Gly Asp Glu Leu His Leu Asp		
	260	265	270
5	Gly Glu Asp Val Ala Met Ala His Ala Asp Ala Leu Asp Asp Phe Asp		
	275	280	285
10	Leu Asp Met Leu Gly Asp Gly Asp Ser Pro Gly Pro Gly Phe Thr Pro		
	290	295	300
	His Asp Ser Ala Pro Tyr Gly Ala Leu Asp Met Ala Asp Phe Glu Phe		
	305	310	315
			320
15	Glu Gln Met Phe Thr Asp Pro Leu Gly Ile Asp Glu Tyr Gly Gly		
	325	330	335

(2) INFORMATION FOR SEQ ID NO:3:

20 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 894 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Herpes Simplex Virus

30 (B) STRAIN: K12, KOS

(C) INDIVIDUAL ISOLATE: tTA_g transactivator

(ix) FEATURE:

(A) NAME/KEY: exon

35 (B) LOCATION: 1..894

(ix) FEATURE:

(A) NAME/KEY: mRNA

(B) LOCATION: 1..894

40 (ix) FEATURE:

(A) NAME/KEY: misc. binding

(B) LOCATION: 1..207

45 (ix) FEATURE:

(A) NAME/KEY: misc. binding

(B) LOCATION: 208..297

(ix) FEATURE:

50 (A) NAME/KEY: CDS

(B) LOCATION: 1..891

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

55	ATG TCT AGA TTA GAT AAA AGT AAA GTG ATT AAC AGC GCA TTA GAG CTG	48
	Met Ser Arg Leu Asp Lys Ser Lys Val Ile Asn Ser Ala Leu Glu Leu	

	1	5	10	15	
	CTT AAT GAG GTC GGA ATC GAA GGT TTA ACA ACC CGT AAA CTC GCC CAG				96
	Leu Asn Glu Val Gly Ile Glu Gly Leu Thr Thr Arg Lys Leu Ala Gln				
	20		25	30	
5	AAG CTA GGT GTA GAG CAG CCT ACA TTG TAT TGG CAT GTA AAA AAT AAG				144
	Lys Leu Gly Val Glu Gln Pro Thr Leu Tyr Trp His Val Lys Asn Lys				
	35	40	45		
	CGG GCT TTG CTC GAC GCC TTA GCC ATT GAG ATG TTA GAT AGG CAC CAT				192
	Arg Ala Leu Leu Asp Ala Leu Ala Ile Glu Met Leu Asp Arg His His				
10	50	55	60		
	ACT CAC TTT TGC CCT TTA GAA GGG GAA AGC TGG CAA GAT TTT TTA CGT				240
	Thr His Phe Cys Pro Leu Glu Gly Glu Ser Trp Gln Asp Phe Leu Arg				
	65	70	75	80	
15	AAT AAC GCT AAA AGT TTT AGA TGT GCT TTA CTA AGT CAT CGC GAT GGA				288
	Asn Asn Ala Lys Ser Phe Arg Cys Ala Leu Leu Ser His Arg Asp Gly				
	85	90	95		
	GCA AAA GTA CAT TTA GGT ACA CGG CCT ACA GAA AAA CAG TAT GAA ACT				336
	Ala Lys Val His Leu Gly Thr Arg Pro Thr Glu Lys Gln Tyr Glu Thr				
	100	105	110		
20	CTC GAA AAT CAA TTA GCC TTT TTA TGC CAA CAA GGT TTT TCA CTA GAG				384
	Leu Glu Asn Gln Leu Ala Phe Leu Cys Gln Gln Gly Phe Ser Leu Glu				
	115	120	125		
	AAT GCA TTA TAT GCA CTC AGC GCT GTG GGG CAT TTT ACT TTA GGT TGC				432
	Asn Ala Leu Tyr Ala Leu Ser Ala Val Gly His Phe Thr Leu Gly Cys				
25	130	135	140		
	GTA TTG GAA GAT CAA GAG CAT CAA GTC GCT AAA GAA GAA AGG GAA ACA				480
	Val Leu Glu Asp Gln Glu His Gln Val Ala Lys Glu Glu Arg Glu Thr				
	145	150	155	160	
30	CCT ACT ACT GAT AGT ATG CCG CCA TTA TTA CGA CAA GCT ATC GAA TTA				528
	Pro Thr Thr Asp Ser Met Pro Pro Leu Leu Arg Gln Ala Ile Glu Leu				
	165	170	175		
	TTT GAT CAC CAA GGT GCA GAG CCA GCC TTC TTA TTC GGC CTT GAA TTG				576
	Phe Asp His Gln Gly Ala Glu Pro Ala Phe Leu Phe Gly Leu Glu Leu				
	180	185	190		
35	ATC ATA TGC GGA TTA GAA AAA CAA CTT AAA TGT GAA AGT GGG TCT GAT				624
	Ile Ile Cys Gly Leu Glu Lys Gln Leu Lys Cys Glu Ser Gly Ser Asp				
	195	200	205		
	CCA TCG ATA CAC ACG CGC AGA CTG TCG ACG GCC CCC CCG ACC GAT GTC				672
	Pro Ser Ile His Thr Arg Arg Leu Ser Thr Ala Pro Pro Thr Asp Val				
40	210	215	220		
	AGC CTG GGG GAC GAG CTC CAC TTA GAC GGC GAG GAC GTG GCG ATG GCG				720
	Ser Leu Gly Asp Glu Leu His Leu Asp Gly Glu Asp Val Ala Met Ala				
	225	230	235	240	
	CAT GCC GAC GCG CTA GAC GAT TTC GAT CTG GAC ATG TTG GGG GAC GGG				768

His Ala Asp Ala Leu Asp Asp Phe Asp Leu Asp Met Leu Gly Asp Gly
245 250 255

5 GAT TCC CCG GGT CCG GGA TTT ACC CCC CAC GAC TCC GCC CCC TAC GGC 816
Asp Ser Pro Gly Pro Gly Phe Thr Pro His Asp Ser Ala Pro Tyr Gly
260 265 270

GCT CTG GAT ATG GCC GAC TTC GAG TTT GAG CAG ATG TTT ACC GAT GCC 864
Ala Leu Asp Met Ala Asp Phe Glu Phe Glu Gln Met Phe Thr Asp Ala
275 280 285

10 CTT GGA ATT GAC GAG TAC GGT GGG TTC TAG 894
Leu Gly Ile Asp Glu Tyr Gly Gly Phe
290 295

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 297 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

20 Met Ser Arg Leu Asp Lys Ser Lys Val Ile Asn Ser Ala Leu Glu Leu
1 5 10 15
Leu Asn Glu Val Gly Ile Glu Gly Leu Thr Thr Arg Lys Leu Ala Gln
20 25 30
25 Lys Leu Gly Val Glu Gln Pro Thr Leu Tyr Trp His Val Lys Asn Lys
35 40 45
Arg Ala Leu Leu Asp Ala Leu Ala Ile Glu Met Leu Asp Arg His His
50 55 60
30 Thr His Phe Cys Pro Leu Glu Gly Glu Ser Trp Gln Asp Phe Leu Arg
65 70 75 80
Asn Asn Ala Lys Ser Phe Arg Cys Ala Leu Leu Ser His Arg Asp Gly
35 85 90 95
Ala Lys Val His Leu Gly Thr Arg Pro Thr Glu Lys Gln Tyr Glu Thr
100 105 110
40 Leu Glu Asn Gln Leu Ala Phe Leu Cys Gln Gln Gly Phe Ser Leu Glu
115 120 125
Asn Ala Leu Tyr Ala Leu Ser Ala Val Gly His Phe Thr Leu Gly Cys
130 135 140
45 Val Leu Glu Asp Gln Glu His Gln Val Ala Lys Glu Glu Arg Glu Thr
145 150 155 160
Pro Thr Thr Asp Ser Met Pro Pro Leu Leu Arg Gln Ala Ile Glu Leu
50 165 170 175

Phe Asp His Gln Gly Ala Glu Pro Ala Phe Leu Phe Gly Leu Glu Leu
180 185 190

5 Ile Ile Cys Gly Leu Glu Lys Gln Leu Lys Cys Glu Ser Gly Ser Asp
195 200 205

Pro Ser Ile His Thr Arg Arg Leu Ser Thr Ala Pro Pro Thr Asp Val
210 215 220

10 Ser Leu Gly Asp Glu Leu His Leu Asp Gly Glu Asp Val Ala Met Ala
225 230 235 240

His Ala Asp Ala Leu Asp Asp Phe Asp Leu Asp Met Leu Gly Asp Gly
15 245 250 255

Asp Ser Pro Gly Pro Gly Phe Thr Pro His Asp Ser Ala Pro Tyr Gly
260 265 270

20 Ala Leu Asp Met Ala Asp Phe Glu Phe Glu Gln Met Phe Thr Asp Ala
275 280 285

Leu Gly Ile Asp Glu Tyr Gly Gly Phe
25 290 295

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 450 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Human cytomegalovirus
- (B) STRAIN: K12, Towne

(ix) FEATURE:

- (A) NAME/KEY: mRNA
- (B) LOCATION: 382..450

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

45 GAATTCCTCG AGTTTACCAC TCCCTATCAG TGATAGAGAA AAGTGAAAGT CGAGTTTACC 60

ACTCCCTATC AGTGATAGAG AAAAGTGAAA GTCGAGTTTA CCACTCCCTA TCAGTGATAG 120

AGAAAAGTGA AAGTCGAGTT TACCACTCCC TATCAGTGAT AGAGAAAAGT GAAAGTCGAG 180

TTTACCACTC CCTATCAGTG ATAGAGAAAA GTGAAAGTCG AGTTTACCAC TCCCTATCAG 240

TGATAGAGAA AAGTGAAAGT CGAGTTTACC ACTCCCTATC AGTGATAGAG AAAAGTGAAA 300

50 GTCGAGCTCG GTACCCGGGT CGAGTAGGCG TGTACGGTGG GAGGCCTATA TAAGCAGAGC 360

TCGTTTAGTG AACCGTCAGA TCGCCTGGAG ACGCCATCCA CGCTGTTTTG ACCTCCATAG 420
AAGACACCGG GACCGATCCA GCCTCCGCGG 450

(2) INFORMATION FOR SEQ ID NO:6:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 450 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
- 10 (ii) MOLECULE TYPE: DNA (genomic)
- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Human cytomegalovirus
(B) STRAIN: Towne
- 15 (ix) FEATURE:
(A) NAME/KEY: mRNA
(B) LOCATION: 382..450

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GAATTCCTCG ACCCGGGTAC CGAGCTCGAC TTCTACTTTT CTCTATCACT GATAGGGAGT 60
GGTAAACTCG ACTTTCACCTT TTCTCTATCA CTGATAGGGA GTGGTAAACT CGACTTTCAC 120
TTTTCTCTAT CACTGATAGG GAGTGGTAAA CTCGACTTTC ACTTTTCTCT ATCACTGATA 180
GGGAGTGGTA AACTCGACTT TCACTTTTCT CTATCACTGA TAGGGAGTGG TAAACTCGAC 240
25 TTCTACTTTT CTCTATCACT GATAGGGAGT GGTAAACTCG ACTTTCACCTT TTCTCTATCA 300
CTGATAGGGA GTGGTAAACT CGAGTAGGCG TGTACGGTGG GAGGCCTATA TAAGCAGAGC 360
TCGTTTAGTG AACCGTCAGA TCGCCTGGAG ACGCCATCCA CGCTGTTTTG ACCTCCATAG 420
AAGACACCGG GACCGATCCA GCCTCCGCGG 450

(2) INFORMATION FOR SEQ ID NO:7:

- 30 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 398 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
- 35 (ii) MOLECULE TYPE: DNA (genomic)
- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Herpes Simplex Virus
(B) STRAIN: KOS
- 40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GAGCTCGACT TTCACTTTTC TCTATCACTG ATAGGGAGTG GTAAACTCGA CTTTCACTTT 60

TCTCTATCAC TGATAGGGAG TGGTAAACTC GACTTTCACT TTTCTCTATC ACTGATAGGG 120
AGTGGTAAAC TCGACTTTCA CTTTTCTCTA TCACTGATAG GGAGTGGTAA ACTCGACTTT 180
CACTTTTCTC TATCACTGAT AGGGAGTGGT AAACTCGACT TTCACTTTTC TCTATCACTG 240
ATAGGGAGTG GTAAACTCGA CTTTCACTTT TCTCTATCAC TGATAGGGAG TGGTAAACTC 300
5 GAGATCCGGC GAATTCTGAAC ACGCAGATGC AGTCGGGGCG GCGCGGTCCG AGGTCCACTT 360
CGCATATTAA GGTGACGCGT GTGGCCTCGA ACACCGAG 398

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 6244 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

15 (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Human cytomegalovirus
(B) STRAIN: Towne (hCMV)

(vii) IMMEDIATE SOURCE:

- (B) CLONE: pUHD BGR3

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CTCGAGTTTA CCACTCCCTA TCAGTGATAG AGAAAAGTGA AAGTCGAGTT TACCACTCCC 60
TATCAGTGAT AGAGAAAAGT GAAAGTCGAG TTTACCACTC CCTATCAGTG ATAGAGAAAA 120
GTGAAAGTCG AGTTTACCAC TCCCTATCAG TGATAGAGAA AAGTGAAAGT CGAGTTTACC 180
ACTCCCTATC AGTGATAGAG AAAAGTGAAA GTCGAGTTTA CCACTCCCTA TCAGTGATAG 240
25 AGAAAAGTGA AAGTCGAGTT TACCACTCCC TATCAGTGAT AGAGAAAAGT GAAAGTCGAG 300
CTCGGTACCC GGGTCGAGTA GCGGTGTACG GTGGGAGGCC TATATAAGCA GAGCTCGTTT 360
AGTGAACCGT CAGATCGCCT GGAGACGCCA TCCACGCTGT TTTGACCTCC ATAGAAGACA 420
CCGGGACCGA TCCAGCCTCC GCGGCCCCGA ATTCGAGCTC GGTACCGGGC CCCCCCTCGA 480
GGTCGACGGT ATCGATAAGC TTGATATCGA ATTCCAGGAG GTGGAGATCC GCGGGTCCAG 540
30 CCAAACCCCA CACCCATTTT CTCCTCCCTC TGCCCCTATA TCCCGGCACC CCCTCCTCCT 600
AGCCCTTTCC CTCCTCCCGA GAGACGGGGG AGGAGAAAAG GGGAGTTTCA GTCGACATGA 660
CTGAGCTGAA GGCAAAGGAA CCTCGGGCTC CCCACGTGGC GGGGGGCGCG CCCTCCCCCA 720
CCGAGGTCGG ATCCCAGCTC CTGGGTGCGC CGGACCCTGG CCCCTTCCAG GGGAGCCAGA 780
CCTCAGAGGC CTCGTCTGTA GTCTCCGCCA TCCCCTCTC CCTGGACGGG TTGCTCTTCC 840

	CCCCGGCCCTG TCAGGGGCAG AACCCCCCAG ACGGGAAGAC GCAGGACCCA CCGTCGTTGT	900
	CAGACGTGGA GGGCGCATTT CCTGGAGTCG AAGCCCCGGA GGGGGCAGGA GACAGCAGCT	960
	CGAGACCTCC AGAAAAGGAC AGCGGCCTGC TGGACAGTGT CCTCGACACG CTCCTGGCGC	1020
	CCTCGGTGCC CGGGCAGAGC CACGCCAGCC CTGCCACCTG CGAGGCCATC AGCCCGTGGT	1080
5	GCCTGTTTGG CCCCACCTT CCCGAAGACC CCCGGGCTGC CCCCCTACC AAAGGGGTGT	1140
	TGGCCCCGCT CATGAGCCGA CCCGAGGACA AGGCAGGCGA CAGCTCTGGG ACGGCAGCGG	1200
	CCCACAAGGT GCTGCCCAGG GGAAGTGTAC CATCCAGGCA GCTGCTGCTC CCCTCCTCTG	1260
	GGAGCCCTCA CTGGCCGGCA GTGAAGCCAT CCCCAGAGCC CGCTGCGGTG CAGGTAGACG	1320
	AGGAGGACAG CTCCGAATCC GAGGGCACCG TGGGCCCCGCT CCTGAAGGGC CAACCTCGGG	1380
10	CACTGGGAGG CACGGCGGCC GGAGGAGGAG CTGCCCCCGT CGCGTCTGGA GCGGCCGCGAG	1440
	GAGGCGTCGC CCTTGTCCCC AAGGAAGATT CTCGTTTCTC GGCGCCCAGG GTCTCCTTGG	1500
	CGGAGCAGGA CGCGCCGGTG GCGCCTGGGC GCTCCCCGCT GGCCACCTCG GTGGTGGATT	1560
	TCATCCACGT GCCCATCCTG CCTCTCAACC ACGCTTTCTT GGCCACCCGC ACCAGGCAGC	1620
	TGCTGGAGGG GGAGAGCTAC GACGGCGGGG CCGCGGCCGC CAGCCCCTTC GTCCCGCAGC	1680
15	GGGGCTCCCC CTCTGCCTCG TCCACCCCTG TGGCGGGCGG CGACTTCCCC GACTGCACCT	1740
	ACCCGCCCCG CGCCGAGCCC AAAGATGACG CGTTCCCCCT CTACGGCGAC TTCCAGCCGC	1800
	CCGCCCTCAA GATAAAGGAG GAGGAAGAAG CCGCCGAGGC CGCGGCGCGC TCCCCGCGTA	1860
	CGTACCTGGT GGCTGGTGCA AACCCCGCCG CCTTCCCGGA CTTCCAGCTG GCAGCGCCGC	1920
	CGCCACCCTC GCTGCCGCCT CGAGTGCCCT CGTCCAGACC CGGGGAAGCG GCGGTGGCGG	1980
20	CCTCCCCAGG CAGTGCCTCC GTCTCCTCCT CGTCCTCGTC GGGGTGACC CTGGAGTGCA	2040
	TCCTGTACAA GGCAGAAGGC GCGCCGCCCC AGCAGGGCCC CTTGCGCCG CTGCCCTGCA	2100
	AGCCTCCGGG CGCCGGCGCC TGCCTGCTCC CGCGGGACGG CCTGCCCTCC ACCTCCGCCT	2160
	CGGGCGCAGC CGCCGGGGCC GCCCCTGCGC TCTACCCGAC GCTCGGCCTC AACGGACTCC	2220
	CGCAACTCGG CTACCAGGCC GCCGTGCTCA AGGAGGGCCT GCCGCAGGTC TACACGCCCT	2280
25	ATCTCAACTA CCTGAGGCCG GATTGAGAAG CCAGTCAGAG CCCACAGTAC AGCTTCGAGT	2340
	CACTACCTCA GAAGATTTGT TTGATCTGTG GGGATGAAGC ATCAGGCTGT CATTATGGTG	2400
	TCCTCACCTG TGGGAGCTGT AAGGTCTTCT TTAAAAGGGC AATGGAAGGG CAGCATAACT	2460
	ATTTATGTGC TGGAAGAAAT GACTGCATTG TTGATAAAAT CCGCAGGAAA AACTGCCCGG	2520
	CGTGTCGCCT TAGAAAGTGC TGTCAAGCTG GCATGGTCCT TGGAGGGCGA AAGTTTAAAA	2580
30	AGTTCAATAA AGTCAGAGTC ATGAGAGCAC TCGATGCTGT TGCTCTCCCA CAGCCAGTGG	2640

	GCATTCCAAA TGAAAGCCAA CGAATCACTT TTTCTCCAAG TCAAGAGATA CAGTTAATTC	2700
	CCCCTCTAAT CAACCTGTTA ATGAGCATTG AACCAGATGT GATCTATGCA GGACATGACA	2760
	ACACAAAGCC TGATACCTCC AGTTCTTTGC TGACGAGTCT TAATCAACTA GGCGAGCGGC	2820
	AACTTCTTTC AGTGGTAAAA TGGTCCAAAT CTCTTCAGG TTTTCGAAAC TTACATATTG	2880
5	ATGACCAGAT AACTCTCATC CAGTATTCTT GGATGAGTTT AATGGTATTT GGACTAGGAT	2940
	GGAGATCCTA CAAACATGTC AGTGGGCAGA TGCTGTATTT TGCACCTGAT CTAATATTAA	3000
	ATGAACAGCG GATGAAAGAA TCATCATTCT ATTCATATG CCTTACCATG TGGCAGATAC	3060
	CGCAGGAGTT TGTCAAGCTT CAAGTTAGCC AAGAAGAGTT CCTCTGCATG AAAGTATTAC	3120
	TACTTCTTAA TACAATTCCT TTGGAAGGAC TAAGAAGTCA AAGCCAGTTT GAAGAGATGA	3180
10	GATCAAGCTA CATTAGAGAG CTCATCAAGG CAATTGGTTT GAGGCAAAAA GGAGTTGTTT	3240
	CCAGCTCACA GCGTTTCTAT CAGCTCACAA AACTTCTTGA TAACTTGCAT GATCTTGTC	3300
	AACAACCTCA CCTGTACTGC CTGAATACAT TTATCCAGTC CCGGGCGCTG AGTGTTGAAT	3360
	TTCCAGAAAT GATGTCTGAA GTTATTGCTG CACAGTTACC CAAGATATTG GCAGGGATGG	3420
	TGAAACCACT TCTCTTTCAT AAAAAGTGAA TGTCAATTAT TTTTCAAAGA ATTAAGTGTT	3480
15	GTGGTATGTC TTTCGTTTTG GTCAGGATTA TGACGTCTCG AGTTTTTATA ATATTCTGAA	3540
	AGGGAATTCC TGCAGCCCGG GGGATCCACT AGTTCTAGAG GATCCAGACA TGATAAGATA	3600
	CATTGATGAG TTTGGACAAA CCACAACCTAG AATGCAGTGA AAAAAATGCT TTATTTGTGA	3660
	AATTTGTGAT GCTATTGCTT TATTTGTAAC CATTATAAGC TGCAATAAAC AAGTTAACAA	3720
	CAACAATTGC ATTCATTTTA TGTTTCAGGT TCAGGGGGAG GTGTGGGAGG TTTTTTAAAG	3780
20	CAAGTAAAAC CTCTACAAAT GTGGTATGGC TGATTATGAT CCTGCAAGCC TCGTCGCTCTG	3840
	GCCGGACCAC GCTATCTGTG CAAGGTCCCC GGACGCGCGC TCCATGAGCA GAGCGCCCGC	3900
	CGCCGAGGCA AGACTCGGGC GGCGCCCTGC CCGTCCCACC AGGTCAACAG GCGGTAACCG	3960
	GCCTCTTCAT CGGGAATGCG CGCGACCTTC AGCATCGCCG GCATGTCCCC TGGCGGACGG	4020
	GAAGTATCAG CTCGACCAAG CTTGGCGAGA TTTTCAGGAG CTAAGGAAGC TAAAATGGAG	4080
25	AAAAAAATCA CTGGATATAC CACCGTTGAT ATATCCCAAT GGCATCGTAA AGAACATTTT	4140
	GAGGCATTTT AGTCAGTTGC TCAATGTACC TATAACCAGA CCGTTCAGCT GCATTAATGA	4200
	ATCGGCCAAC GCGCGGGGAG AGGCGGTTTG CGTATTGGGC GCTCTTCCGC TTCCTCGCTC	4260
	ACTGACTCGC TGCCTCGGT CGTTCGGCTG CGGCGAGCGG TATCAGCTCA CTCAAAGGCG	4320
	GTAATACGGT TATCCACAGA ATCAGGGGAT AACGCAGGAA AGAACATGTG AGCAAAAGGC	4380
30	CAGCAAAAGG CCAGGAACCG TAAAAGGCC GCGTTGCTGG CGTTTTTCCA TAGGCTCCGC	4440

	CCCCCTGACG AGCATCACAA AAATCGACGC TCAAGTCAGA GGTGGCGAAA CCCGACAGGA	4500
	CTATAAAGAT ACCAGGCGTT TCCCCCTGGA AGCTCCCTCG TGCGCTCTCC TGTTCCGACC	4560
	CTGCCGCTTA CCGGATACCT GTCCGCCTTT CTCCCTTCGG GAAGCGTGGC GCTTTCTCAA	4620
	TGCTCACGCT GTAGGTATCT CAGTTCGGTG TAGGTCGTTT GCTCCAAGCT GGGCTGTGTG	4680
5	CACGAACCCC CCGTTCAGCC CGACCGCTGC GCCTTATCCG GTAACATATCG TCTTGAGTCC	4740
	AACCCGGTAA GACACGACTT ATCGCCACTG GCAGCAGCCA CTGGTAACAG GATTAGCAGA	4800
	GCGAGGTATG TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACTA CGGCTACACT	4860
	AGAAGGACAG TATTTGGTAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG AAAAAGAGTT	4920
	GGTAGCTCTT GATCCGGCAA ACAAACCACC GCTGGTAGCG GTGGTTTTTT TGTTTGCAAG	4980
10	CAGCAGATTA CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG	5040
	TCTGACGCTC AGTGGAACGA AAATCACGT TAAGGGATTT TGGTCATGAG ATTATCAAAA	5100
	AGGATCTTCA CCTAGATCCT TTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA	5160
	TATGAGTAAA CTTGGTCTGA CAGTTACCAA TGCTTAATCA GTGAGGCACC TATCTCAGCG	5220
	ATCTGTCTAT TTCGTTTCATC CATAGTTGCC TGACTCCCCG TCGTGTAGAT AACTACGATA	5280
15	CGGGAGGGCT TACCATCTGG CCCAGTGCT GCAATGATAC CGCGAGACCC ACGCTCACCG	5340
	GCTCCAGATT TATCAGCAAT AAACCAGCCA GCCGGAAGGG CCGAGCGCAG AAGTGGTCCT	5400
	GCAACTTTAT CCGCCTCCAT CCAGTCTATT AATTGTTGCC GGAAGCTAG AGTAAGTAGT	5460
	TCGCCAGTTA ATAGTTTGCG CAACGTTGTT GCCATTGCTA CAGGCATCGT GGTGTCACGC	5520
	TCGTGCTTTG GTATGGCTTC ATTCAGCTCC GGTTCCCAAC GATCAAGGCG AGTTACATGA	5580
20	TCCCCCATGT TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT	5640
	AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC TCTTACTGTC	5700
	ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA	5760
	TAGTGTATGC GGCGACCGAG TTGCTCTTGC CCGGCGTCAA TACGGGATAA TACCGCGCCA	5820
	CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA	5880
25	AGGATCTTAC CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT	5940
	TCAGCATCTT TTAATTTTAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG GCAAAATGCC	6000
	GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTCAA	6060
	TATTATTGAA GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT	6120
	TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC	6180
30	TAAGAAACCA TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT	6240

CGTC

6244

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4963 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Human cytomegalovirus

(vii) IMMEDIATE SOURCE:

- (B) CLONE: pUHD BGR4

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CTCGAGTTTA CCACTCCCTA TCAGTGATAG AGAAAAGTGA AAGTCGAGTT TACCACTCCC	60
TATCAGTGAT AGAGAAAAGT GAAAGTCGAG TTTACCACTC CCTATCAGTG ATAGAGAAAA	120
GTGAAAGTCG AGTTTACCAC TCCCTATCAG TGATAGAGAA AAGTGAAAGT CGAGTTTACC	180
ACTCCCTATC AGTGATAGAG AAAAGTGAAA GTCGAGTTTA CCACTCCCTA TCAGTGATAG	240
AGAAAAGTGA AAGTCGAGTT TACCACTCCC TATCAGTGAT AGAGAAAAGT GAAAGTCGAG	300
CTCGGTACCC GGGTCGAGTA GGCCTGTACG GTGGGAGGCC TATATAAGCA GAGCTCGTTT	360
AGTGAACCGT CAGATCGCCT GGAGACGCCA TCCACGCTGT TTTGACCTCC ATAGAAGACA	420
CCGGGACCGA TCCAGCCTCC GCGGCCCCGA ATTCCGGCCA CGACCATGAC CATGACCCTC	480
CACACCAAAG CATCTGGGAT GGCCCTACTG CATCAGATCC AAGGGAACGA GCTGGAGCCC	540
CTGAACCGTC CGCAGCTCAA GATCCCCCTG GAGCGGCCCC TGGGCGAGGT GTACCTGGAC	600
AGCAGCAAGC CCGCCGTGTA CAACTACCCC GAGGGCGCCG CCTACGAGTT CAACGCCGCG	660
GCCGCCGCCA ACGCGCAGGT CTACGGTCAG ACCGGCCTCC CCTACGGCCC CGGGTCTGAG	720
GCTGCGGCGT TCGGCTCCAA CGGCCTGGGG GGTTCCTCCC CACTCAACAG CGTGTCTCCG	780
AGCCCGCTGA TGCTACTGCA CCCGCCGCCG CAGCTGTGCG CTTTCCTGCA GCCCCACGGC	840
CAGCAGGTGC CCTACTACCT GGAGAACGAG CCCAGCGGCT ACACGGTGCG CGAGGCCGGC	900
CCGCCGGCAT TCTACAGGCC AAATTCAGAT AATCGACGCC AGGGTGGCAG AGAAAGATTG	960
GCCAGTACCA ATGACAAGGG AAGTATGGCT ATGGAATCTG CCAAGGAGAC TCGCTACTGT	1020
GCAGTGTGCA ATGACTATGC TTCAGGCTAC CATTATGGAG TCTGGTCTCTG TGAGGGCTGC	1080
AAGGCCTTCT TCAAGAGAAG TATTCAAGGA CATAACGACT ATATGTGTCC AGCCACCAAC	1140
CAGTGCACCA TTGATAAAAA CAGGAGGAAG AGCTGCCAGG CCTGCCGGCT CCGCAAATGC	1200

	TACGAAGTGG GAATGATGAA AGGTGGGATA CGAAAAGACC GAAGAGGAGG GAGAATGTTG	1260
	AAACACAAGC GCCAGAGAGA TGATGGGGAG GGCAGGGGTG AAGTGGGGTC TGCTGGAGAC	1320
	ATGAGAGCTG CCAACCTTTG GCCAAGCCCG CTCATGATCA AACGCTCTAA GAAGAACAGC	1380
	CTGGCCTTGT CCCTGACGGC CGACCAGATG GTCATGGCCT TGTTGGATGC TGAGCCCCCC	1440
5	ATACTCTATT CCGAGTATGA TCCTACCAGA CCCTTCAGTG AAGCTTCGAT GATGGGCTTA	1500
	CTGACCAACC TGGCAGACAG GGAGCTGGTT CACATGATCA ACTGGGCGAA GAGGGTGCCA	1560
	GGCTTTGTGG ATTTGACCCT CCATGATCAG GTCCACCTTC TAGAATGTGC CTGGCTAGAG	1620
	ATCCTGATGA TTGGTCTCGT CTGGCGCTCC ATGGAGCACC CAGTGAAGCT ACTGTTTGCT	1680
	CCTAACTTGC TCTTGACAG GAACCAGGGA AAATGTGTAG AGGGCATGGT GGAGATCTTC	1740
10	GACATGCTGC TGGCTACATC ATCTCGGTTT CGCATGATGA ATCTGCAGGG AGAGGAGTTT	1800
	GTGTGCCTCA AATCTATTAT TTTGCTTAAT TCTGGAGTGT ACACATTTCT GTCCAGCACC	1860
	CTGAAGTCTC TGGAAGAGAA GGACCATATC CACCGAGTCC TGGACAAGAT CACAGACACT	1920
	TTGATCCACC TGATGGCCAA GGCAGGCCTG ACCCTGCAGC AGCAGCACCA GCGGCTGGCC	1980
	CAGCTCCTCC TCATCCTCTC CCACATCAGG CACATGAGTA ACAAAGGCAT GGAGCATCTG	2040
15	TACAGCATGA AGTGCAAGAA CGTGGTGCCC CTCTATGACC TGCTGCTGGA GATGCTGGAC	2100
	GCCCACCGCC TACATGCGCC CACTAGCCGT GGAGGGGCAT CCGTGGAGGA GACGGACCAA	2160
	AGCCACTTGG CCACTGCGGG CTCTACTTCA TCGCATTCTT TGCAAAAGTA TTACATCACG	2220
	GGGGAGGCAG AGGGTTTCCC TGCCACAGTC TGAGAGCTCC CTGGCGGAAT TCGAGCTCGG	2280
	TACCCGGGGA TCCTCTAGAG GATCCAGACA TGATAAGATA CATTGATGAG TTTGGACAAA	2340
20	CCACAAC TAG AATGCAGTGA AAAAAATGCT TTATTTGTGA AATTTGTGAT GCTATTGCTT	2400
	TATTTGTAAC CATTATAAGC TGCAATAAAC AAGTTAACAA CAACAATTGC ATTCATTTTA	2460
	TGTTTCAGGT TCAGGGGGAG GTGTGGGAGG TTTTTTAAAG CAAGTAAAAC CTCTACAAAT	2520
	GTGGTATGGC TGATTATGAT CCTGCAAGCC TCGTCGTCTG GCCGGACCAC GCTATCTGTG	2580
	CAAGGTCCCC GGACGCGCGC TCCATGAGCA GAGCGCCCGC CGCCGAGGCA AGACTCGGGC	2640
25	GGCGCCCTGC CCGTCCACAC AGGTCAACAG GCGGTAACCG GCCTCTTCAT CGGGAATGCG	2700
	CGCGACCTTC AGCATCGCCG GCATGTCCCC TGGCGGACGG GAAGTATCAG CTCGACCAAG	2760
	CTTGGCGAGA TTTTCAGGAG CTAAGGAAGC TAAAATGGAG AAAAAATCA CTGGATATAC	2820
	CACCGTTGAT ATATCCCAAT GGCATCGTAA AGAACATTTT GAGGCATTTT AGTCAGTTGC	2880
	TCAATGTACC TATAACCAGA CCGTTCAGCT GCATTAATGA ATCGGCCAAC GCGCGGGGAG	2940
30	AGGCGGTTTG CGTATTGGGC GCTCTTCCGC TTCCTCGCTC ACTGACTCGC TGCGCTCGGT	3000

	CGTTCGGCTG CGGCGAGCGG TATCAGCTCA CTCAAAGGCG GTAATACGGT TATCCACAGA	3060
	ATCAGGGGAT AACGCAGGAA AGAACATGTG AGCAAAAGGC CAGCAAAAGG CCAGGAACCG	3120
	TAAAAAGGCC GCGTTGCTGG CGTTTTTCCA TAGGCTCCGC CCCCCTGACG AGCATCACAA	3180
	AAATCGACGC TCAAGTCAGA GGTGGCGAAA CCCGACAGGA CTATAAAGAT ACCAGGCGTT	3240
5	TCCCCCTGGA AGCTCCCTCG TGCCTCTCC TGTTCCGACC CTGCCGCTTA CCGGATACCT	3300
	GTCCGCCTTT CTCCCTTCGG GAAGCGTGGC GCTTTCTCAA TGCTCACGCT GTAGGTATCT	3360
	CAGTTCGGTG TAGGTCGTTT GCTCCAAGCT GGGCTGTGTG CACGAACCCC CCGTTCAGCC	3420
	CGACCGCTGC GCCTTATCCG GTAACATATCG TCTTGAGTCC AACCCGGTAA GACACGACTT	3480
	ATCGCCACTG GCAGCAGCCA CTGGTAACAG GATTAGCAGA GCGAGGTATG TAGGCGGTGC	3540
10	TACAGAGTTC TTGAAGTGGT GGCCTAACTA CGGCTACACT AGAAGGACAG TATTTGGTAT	3600
	CTGCGCTCTG CTGAAGCCAG TTACCTTCGG AAAAAAGAGTT GGTAGCTCTT GATCCGGCAA	3660
	ACAAACCACC GCTGGTAGCG GTGGTTTTTT TGTTCGCAAG CAGCAGATTA CGCGCAGAAA	3720
	AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA	3780
	AAACTCACGT TAAGGGATTT TGGTCATGAG ATTATCAAAA AGGATCTTCA CCTAGATCCT	3840
15	TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA CTTGGTCTGA	3900
	CAGTTACCAA TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTTCATC	3960
	CATAGTTGCC TGATCCCCGT CGTGTAGATA ACTACGATAC GGGAGGGCTT ACCATCTGGC	4020
	CCCAGTGCTG CAATGATACC GCGAGACCCA CGCTACCGG CTCCAGATTT ATCAGCAATA	4080
	AACCAGCCAG CCGGAAGGGC CGAGCGCAGA AGTGGTCCTG CAACTTTATC CGCCTCCATC	4140
20	CAGTCTATTA ATTGTTGCCG GGAAGCTAGA GTAAGTAGTT CGCCAGTTAA TAGTTTGCGC	4200
	AACGTTGTTG CCATTGCTAC AGGCATCGTG GTGTCACGCT CGTCGTTTGG TATGGCTTCA	4260
	TTCAGCTCCG GTTCCCAACG ATCAAGGCGA GTTACATGAT CCCCCATGTT GTGCAAAAAA	4320
	GCGGTTAGCT CCTTCGGTCC TCCGATCGTT GTCAGAAGTA AGTTGGCCGC AGTGTTATCA	4380
	CTCATGGTTA TGGCAGCACT GCATAATTCT CTTACTGTCA TGCCATCCGT AAGATGCTTT	4440
25	TCTGTGACTG GTGAGTACTC AACCAAGTCA TTCTGAGAAT AGTGTATGCG GCGACCGAGT	4500
	TGCTCTTGCC CGGCGTCAAT ACGGGATAAT ACCGCGCCAC ATAGCAGAAC TTTAAAAGTG	4560
	CTCATCATTG GAAAACGTTT TCCGGGGCGA AAACCTCTCA GGATCTTACCGCTGTTGAGA	4620
	TCCAGTTCGA TGTAACCCAC TCGTGCACCC AACTGATCTT CAGCATCTTT TACTTTACCC	4680
	AGCGTTTCTG GGTGAGCAAA AACAGGAAGG CAAAATGCCG CAAAAAAGGG AATAAGGGCG	4740
30	ACACGGAAAT GTTGAATACT CATACTCTTC CTTTTTCAAT ATTATTGAAG CATTTATCAG	4800

GGTTATTGTC TCATGAGCGG ATACATATTT GAATGTATTT AGAAAAATAA ACAAATAGGG	4860
GTTCCGCGCA CATTTCCCCG AAAAGTGCCA CCTGACGTCT AAGAAACCAT TATTATCATG	4920
ACATTAACCT ATAAAAATAG GCGTATCACG AGGCCCTTTC GTC	4963

5 (2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 42 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TCGAGTTTAC CACTCCCTAT CAGTGATAGA GAAAAGTGAA AG

42